IIAI Open Conference Publication Series
IIAI Letters on Informatics and Interdisciplinary Research
Vol.006, LIIR455
DOI: https://doi.org/10.52731/liir.v006.455

Proposal for a Novel Recipe Generation Approach Using Inverse TF-IDF and Generative AI

Mutsumi Kodama [*], Taishi Nemoto [†]

Abstract

This paper proposes a novel recipe generation system that integrates an inverse term frequency–inverse document frequency (TF-IDF) approach with generative artificial intelligence (AI) to balance creativity with practicality in home cooking. While conventional recipe recommendation systems tend to focus on commonplace recipes or staple ingredients, this study adopted an inverse approach by focusing on low-IDF ingredients and leveraging them as the foundation for generating highly reproducible recipes. Utilizing GPT-4, the system automatically constructs prompts that reflect user preferences and cooking conditions, enabling the generation of context-sensitive recipes. A lightweight web-based demo system was developed, allowing users to input two ingredients and receive original recipe suggestions along with cosine similarity scores in real time. Even when users input seemingly unrelated ingredients, the system provides composite recipe outputs, demonstrating its creative flexibility. This approach uncovers the latent creative potential in common ingredients and suggests broad applicability in the food industry, nutritional planning, and culinary education. By combining generative AI with statistical ingredient analysis, the proposed method offers a new intelligent support model for everyday cooking practices.

***Keywords**:* Artificial Intelligence, Education, Recipe 3.0, LLMs for cooking, nutritional analysis, democratization of culinary innovation, ChatGPT

1 Introduction

Food culture in modern society is more diverse and complex than ever before. With the advancement of globalization and information technology, the choice of ingredients has expanded, and cooking methods and taste preferences have also diversified. In addition, the roles expected of food, such as increasing health consciousness, promoting interest in sustainability, and encouraging consideration for food allergies, have expanded beyond the traditional scope of "nutritional supplementation" and have become important elements in improving quality of life. Home cooking serves as a flexible setting to meet such complex demands; however, creating daily menus, selecting ingredients, and deciding on cooking methods places a high cognitive load. In particular, preparing healthy and satisfying meals within a limited time has become difficult in modern dual-income households, highlighting the need for assistive technologies.

[*] Kagoshima Women's College ,Kagoshima, Japan
[†] University of Kochi, Kochi, Japan

In this context, the automatic generation of cooking recipes using generative artificial intelligence (AI) has attracted considerable attention in recent years. Advances in natural language processing (NLP) using large language models (LLMs) are enabling AI to understand context and make creative combinations that were previously difficult even for humans. Models such as ChatGPT and GPT-4 have achieved excellent results in fields such as everyday conversations, programming tasks, and educational content generation. Applying these technologies to recipe generation is expected to enable the proposal of recipes that meet individual needs.

However, current AI recipe generation methods have several issues. First, the output recipes based on existing recipe data tend to merely imitate existing patterns and lack originality. Second, placing too much emphasis on originality can result in a loss of practicality and reproducibility. Designing new algorithms that achieve a high-level balance between creativity and practicality is necessary to resolve these issues.

This research focuses on the Term Frequency-Inverse Document Frequency (TF-IDF) method, which has been widely used in the field of NLP, and proposes a "reverse TF-IDF approach" based on this reverse idea. In conventional TF-IDF, the rarer a word is in a sentence, the more important it is. However, in this research, we aim to generate creative recipes by combining ingredients commonly kept at home with new combinations, placing more weight on “frequent words = common ingredients.”

2 Purpose

This research aims to overcome the limitations of existing recipe generation systems that rely on existing patterns and build a practical and creative recipe generation model based on common household ingredients. Conventional systems use the TF-IDF method to extract statistical features from existing recipe data, but because the method places weight on rare words, it does not fully evaluate the practicality and reproducibility of home cooking or the everyday value of ingredients. This research introduces a new perspective called inverse TF-IDF and focuses on the strong practicality and ease of preparation of commonly used low-IDF ingredients that appear frequently. In other words, it is assumed that ingredients widely used in daily life are the factors that ensure reproducibility, cooking safety, and taste stability in actual home cooking.

Furthermore, advances in generative AI have made it possible to automatically generate recipe ideas by leveraging vast text generation capabilities. However, whether the generated recipe meets specific user requirements, such as cooking time, taste preferences, and nutritional balance, depends on the prompts entered. This research employs a prompt-generation mechanism to personalize recipes based on user preferences and context and aims to provide creative yet highly practical recipes that are free from conventional fixed patterns by combining low-IDF ingredient information obtained through inverse TF-IDF with user requirements.

The purpose of this research is, first, to use the new concept of inverse TF-IDF to extract the creative potential hidden in the everyday nature of common ingredients, in contrast to the scarcity information traditionally emphasized. Second, by combining inverse TF-IDF with generative AI, we aim to establish a personalized recipe generation model that is easy for users to use and that

balances practicality and innovation in home cooking. Furthermore, we aim to establish a technical framework as a foundation for applications in the food industry, health management, and education, constructing a new recipe generation model with wide-ranging social impact. Based on the above considerations, the purpose of this research is to go beyond the reliance on fixed concepts of conventional recipe generation systems and to realize a system that automatically generates cooking ideas that can be adopted at home and flexibly responds to the diverse needs of users.

3 Background

In modern society, consumers' food awareness is diversifying, and they are now seeking multiple values beyond the scope of simple nutritional supplementation, such as health, environment, culture, and psychological satisfaction. Consumers are not only conscious of nutritional management for maintaining health and preventing disease, but also choose meals that contribute to the realization of each family's preferences, local culture, and even sustainable lifestyles. However, conventional recipe-sharing platforms simply present vast amounts of existing recipe data, making it difficult to reflect the detailed requirements of individual users, such as their health status, cooking environment, and taste preferences. Under these circumstances, home cooking requires not only novelty but also practicality, healthiness, and highly reproducible.

The TF-IDF method, which is widely used in information retrieval and text mining, originally extracts document-specific features by giving high weights to rare words in a document. However, in the context of cooking recipes, frequently occurring ingredients and cooking terms are key to practicality and reproducibility. Conventional TF-IDF methods reduce the presence of these common ingredients, so their basic aspects of home cooking are not fully evaluated. Currently, research on automatic recipe generation using generative AI mainly attempts to propose new recipes by learning the structure and language patterns of existing recipes, but the generated recipes often do not correspond to the user's real cooking environment. In real households, dishes that are highly reproducible using limited ingredients and basic cooking methods are required, so an approach rooted in the context of home cooking is needed.

In this context, this study adopts an inverse TF-IDF approach that shifts the concept of TF-IDF and focuses on common ingredients (low-IDF ingredients). Low-IDF ingredients are easily available in ordinary households and function as elements that increase the reproducibility and safety of cooking. Therefore, reevaluating the everyday nature of low-IDF ingredients is considered extremely useful in expanding the possibilities for generating recipes through new creative combinations while utilizing ingredients that users are already familiar with. Furthermore, large-scale language models, which have recently advanced, are capable of flexible sentence generation according to context. Combining this with the inverse TF-IDF approach, is expected automatically generate novel and practical recipes that go beyond conventional fixed recipe patterns. This research background supports the significance of this study as a new technical approach to balance practicality and creativity in home cooking.

4 Precedent Research

Past research on recipe generation systems has primarily analyzed existing recipe data using statistical or rule-based methods to propose new recipe ideas. One early example is Chef Watson, developed by IBM, which generates recipes by integrating flavor profiles and ingredient interactions modeled by Bayesian and statistical inference. This approach has attracted attention for extracting patterns in the flavor composition of ingredients and cooking methods and presenting new combinations that differ from conventional recipes, but the generated results often merely extend existing recipes.

In recent years, studies have applied large-scale language models such as GPT-4. These methods are capable of generating natural sentences based on knowledge learned from a wide range of text data and have been applied to generate complex linguistic expressions, such as recipes. Models such as RecipeGPT have been fine-tuned to capture recipe-specific contexts and generate new recipes automatically. However, these systems still rely heavily on existing data patterns and face limitations in terms of practicality, particularly in adapting to the ingredients and cooking environments used in real households.

On the other hand, numerous TF-IDF–based recipe recommenders have been proposed. These systems use a method to extract recipes that meet individual needs by calculating the similarity between ingredients in recipe data and those in the user's inventory. However, because conventional TF-IDF places high weight on rare words, the reproducibility and practicality of common ingredients regularly used in home cooking are not fully evaluated, and as a result, the generated recipes tend to be conservative.

Furthermore, as an application of generative AI, attempts have been reported to use general-purpose large-scale language models, such as ChatGPT and LLaMA, to provide personalized recipe suggestions based on user's preferences and health information. Although these systems aim to improve the quality of the generated text by reflecting human interaction and feedback, the overall system lacks novel breakthroughs beyond existing patterns in terms of creativity.

In light of the above background, this study clearly identifies the limitations of previous research and introduces a new approach using reverse TF-IDF, focusing on common ingredients that are widely available in households, in contrast with the conventional approach that focuses on rare words. This method combines conventional statistical methods with natural language generation to leverage both the practicality of familiar ingredients and the creative text generation capabilities of generative AI. It is considered a meaningful attempt to differentiate it from previous research. In addition, the fusion of generative AI and TF-IDF is expected to revolutionize existing methods that have been limited to fixed recipe patterns and provide a theoretical foundation for a new recipe generation model with wide applicability.

5 Proposal System

This research proposes a recipe generation system that breaks away from conventional, fixed recipe provision methods. It integrates an inverse TF-IDF approach that focuses on common ingredients available at home and leverages generative AI technology to create more flexible and personalized recipes.

The system begins by collecting and preprocessing a large-scale recipe corpus and then calculates the TF-IDF by calculating the frequency of ingredients in each recipe. Here, our system emphasizes low-IDF (common) ingredients rather than rare ones, evaluating their actual use in households and reproducibility. We used inverse TF-IDF scores to generate prompts, which form the core of the system.

Specifically, the system uses the extracted list of low-IDF ingredients to automatically generate prompts, which are combined with specific requirements entered by the user (cooking time, taste preferences, nutritional balance, etc.). These prompts are sent to a generative AI (a GPT-based model) that outputs new and creative recipe texts without relying solely on existing recipe data.

The entire system is built on a back-end API server with multiple modules working together. First, the collected recipe data are preprocessed, followed by extraction of low-IDF ingredients and TF-IDF calculation. Then, the prompt generator integrates low-IDF ingredients with user requests information to form an optimal prompt. Furthermore, the generation AI automatically generates recipe candidates based on the given prompt, allowing flexible generation that reflects the user's preferences rather than being limited to conventional fixed patterns. The user interface features an intuitive web/mobile interface, allowing users to easily check the generated results and make fine adjustments or request regeneration as necessary.

By adopting the new inverse TF-IDF approach, the proposed system simultaneously evaluates both the practical and creative values of ingredients used on a daily basis. It overcomes the problems of conventional recipe generation methods, such as excessive reliance on existing data and fixed output, and provides a new foundation for automatically generating recipes that users can easily prepare at home. This allows us to present recipes that can be adopted in home cooking and demonstrate the practicality and innovation of the system, which is expected to be deployed in the food industry, nutritional management, and education fields.

6 Mechanism

The mechanism underlying this research is based on the inverse application of the conventional TF-IDF method and aims to provide a new framework for balancing practicality and creativity in home cooking. Typically, TF-IDF assigns high weights to rare words in a document and is used to extract features specific to that document. However, this system adopts the concept of inverse TF-IDF and focuses on low IDF values, that is, ingredients commonly used in many recipes. This creates a mechanism for reevaluating the universality of ingredients used in ordinary households and emphasizing their practicality and reproducibility in cooking.

Specifically, the TF-IDF score for each ingredient is obtained by calculating the frequency of

occurrence of each ingredient in the entire recipe corpus. Here, common ingredients such as eggs, milk, onions, soy sauce, and sugar, which are extracted as low-IDF ingredients, form the basis of home cooking. For example, words such as “stir-fry,” “soy sauce,” “rice,” and “fried rice” often co-occur in recipe document, so they form a region close to each other in the vector space. As a result, documents in which these words frequently appear have high cosine similarity and are located close to each other. Conversely, words such as “lettuce,” “shabu-shabu,” and “spicy” are used in contexts with considerably different cooking methods and uses, so they are located farther away in the TF-IDF vector space compared to the above examples. These distribution patterns support the theoretical basis that common ingredients with which users are already familiar play a core role in the generation process.

This reverse TF-IDF information is reflected in the subsequent prompt generation process. The system integrates the extracted list of low-IDF ingredients with the specific requirements provided by the user (cooking time, taste preferences, nutritional needs, etc.) and automatically generates a prompt sentence in which the list of low-IDF ingredients is explicitly inserted according to a predefined template. The generated prompt encourages the creation of practical recipe ideas that are not limited to existing patterns and are in line with the actual circumstances of the household. The generative AI takes this prompt as input and uses natural language generation technology to output text that contains information, such as cooking steps, utensils used, and flavor harmony that can be easily reproduced at home. In other words, reverse TF-IDF functions as a new indicator for balancing creative combinations hidden in generality and practical reproducibility by reevaluating the distance between each word in the TF-IDF vector space.

Furthermore, to improve the quality of the generated recipes, this system focuses on the role of frequently occurring fixed expressions and cooking terms in forming the structure of the document, in addition to the reproducibility of low-IDF ingredients. Conventionally, general terms receive low TF-IDF weights, but this system positions these frequently occurring elements to play an important role in supporting procedural clarity and readability. For example, fixed expressions such as "add ~" and "please ~" are essential elements for making the generated recipes easy for users to understand and reproduce in actual cooking. By focusing on this point, the recipes proposed by the generative AI are designed not only to be original but also to have a style and structure that makes cooking at home easy.

In addition, the theoretical significance of reverse TF-IDF is not merely a statistical calculation method but is based on the duality of reinterpreting information retrieval and reevaluating everyday life. First, while conventional TF-IDF focuses on scarce information, reverse TF-IDF focuses on commonality and universality, deriving new value from the "common" elements behind home cooking. Second, it has a practical effect in that common ingredients found in daily life provide stable reproducibility and a sense of security in actual cooking. Third, from the perspective that frequently occurring linguistic expressions form the structural foundation of a document, placing emphasis on this information makes the entire text of the generated recipe consistent and the format understandable for users. This multifaceted theoretical framework maximizes the flexible text generation capabilities of generative AI and serves as a foundation for achieving both practicality and creativity in the field of home cooking.

Furthermore, generative AI technology enables linguistic nuances and creative expression that could not be captured by conventional rule-based methods. The fusion of prompt design based on user requests and low-IDF ingredient information based on inverse TF-IDF drives generative

AI to create new recipe texts without being trapped by the existing context. As a result, the generated recipes are practical enough for users to use in actual cooking situations and also have novelty and creativity not been seen before. As part of practical innovation in home cooking, this mechanism will contribute to opening up possibilities for applications in the food industry, nutritional management, and education.

7 Demo System

We describe a web-based demo system that demonstrates its effectiveness. This system is an interactive GPT-4 application that generates creative and practical recipes based on ingredient pairs derived using the inverse TF-IDF approach. The user interface has an intuitive design, and by inputting two ingredients, users can simultaneously obtain recipes based on the combination and vector similarity information.
This demo system is built with a backend configuration using Flask and an HTML/CSS frontend, and has a lightweight design that is easy to deploy and run. Using the OpenAI API, the system builds prompts and returns results in real time according to user input, instantly reflecting the recipe generation results on the web page. In addition, the ingredient vectors are trained on a proprietary recipe corpus, enabling similarity evaluation based on the distribution and semantic proximity of frequently occurring ingredients.

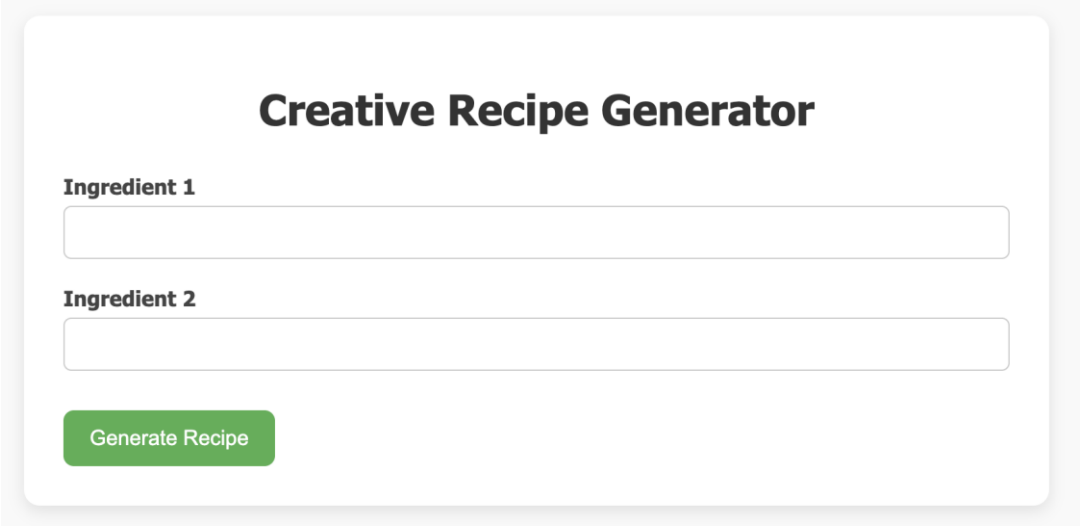

Figure 1: Demo of two ingredient input

The most distinctive feature of this demo system is that it does not simply match existing recipes to the two ingredients selected by the user but instead uses the contextual creativity of a language model to generate a unique recipe. The generated recipe includes the name of the dish, the flavor direction (e.g., sweet and spicy, light, rich), necessary ingredients (up to five), cooking instructions, etc., and is presented at a level that can be reproduced at home.

In addition, the cosine similarity between two ingredients is displayed as a numerical value. This is calculated based on the food vector space learned by Word2Vec, and the higher the similarity, the closer the relationship in terms of taste and usage context. Conversely, when the similarity is low, highly creative recipes are expected to be generated in unexpected combinations. Using

similarity scores as a guide, users can try new cooking suggestions that are not bound by existing cooking contexts. Through this system, users can easily obtain creative cooking ideas for low-IDF food combinations and experience adding new ideas to everyday home cooking. Furthermore, by introducing features such as saving, regenerating, and comparing the obtained recipes, this demo can also serve as a foundation for expanding the educational and practical possibilities. In the future, we are considering expanding this interface and integrating modules for nutritional analysis, allergy information, cost estimation.

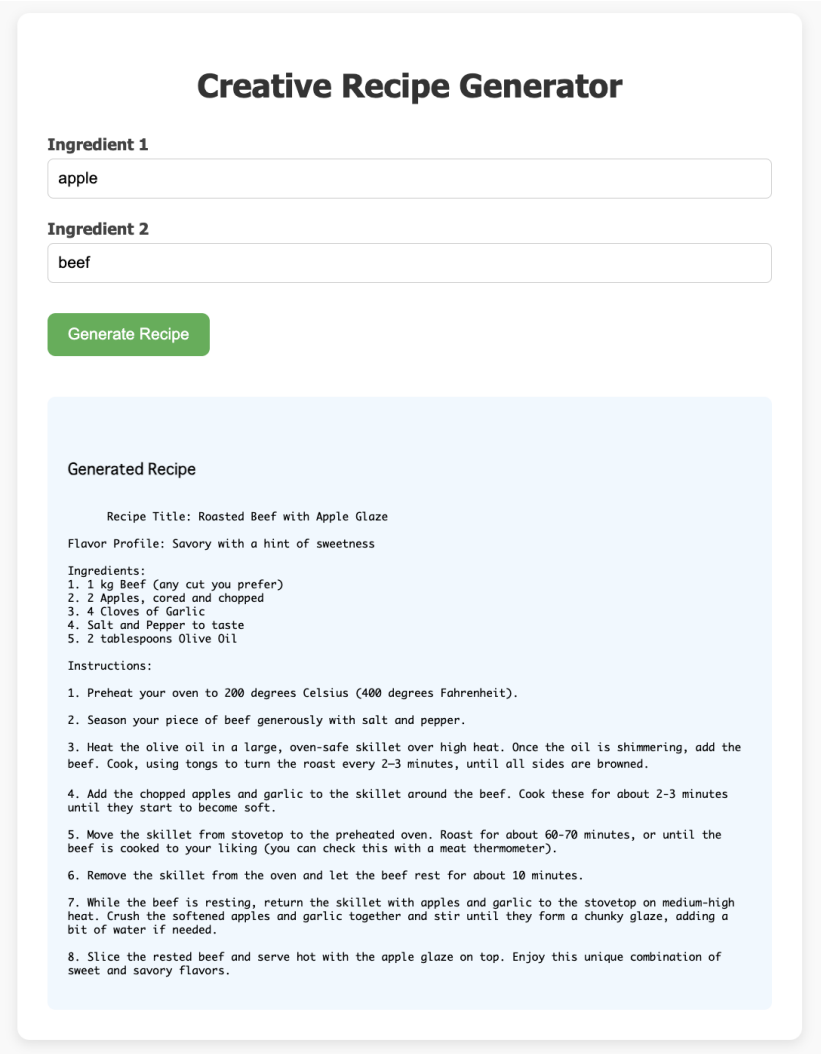

Figure 2: Demo Image of generated recipe

8 Conclusion

In this paper, we propose a new recipe generation system that combines the practicality and creativity of common ingredients in home cooking by combining generative AI and inverse TF-IDF approaches to overcome the reliance on fixed patterns in conventional systems. Conventional TF-IDF extracts document-specific features by emphasizing rare words, but this study attempts to reverse this. We used a GPT-based model with prompts derived from inverse TF-IDF scores, which enabled flexible and highly novel recipe text generation without relying on existing data

patterns. This system is a newly proposed model to replace conventional fixed recipe-posting services as a method for users to easily obtain innovative cooking ideas from common ingredients in home cooking. Extraction of low-IDF ingredients by reverse TF-IDF and generation of prompts reflecting them bring a new perspective to existing techniques of information retrieval and lead to a reevaluation of creative elements hidden in everyday life.

Of course, creating perfect and satisfying recipes without human intervention remains difficult for AI, and elements such as taste, cultural context, and physical sensations are still left to human judgment. However, this system can accelerate and support the trial-and-error cycle of ideation, combination, and verification in recipe development. In other words, by linking human creativity with AI's exploratory capabilities, it has the potential to become a new intellectual support platform for recipe development. Future studies will include integrating nutritional analysis modules to enable the generation of recipes that can be used in actual cooking situations. Considering the possibility of applying this system in the food industry, nutritional management, and even in the field of education is also necessary. These results overcome the limitations of conventional methods and offer promising avenues for practical deployment in home cooking, industry, and education.

References

[1] Kanwipa Lertsumruaypun, Satoshi Nakamura, Chiemi Watanabe "Association of Onomatopoeia and Recipe Terminology in Recipe Search Systems" ,Information Processing Society of Japan Research Report, IPSJ SIG Technical Report, 2010

[2] Yasuyuki Akazawa, Hisashi Miyamori, "Proposal of a Cooking Recipe Search System Considering Refrigerator Ingredients" ,DEIM Forum 2012

[3] Kyosuke Ikeijiri,Yuichi Sei, Hiroyuki Nakagawa, Yasuyuki Tahara, Akihiko Ohsuga, " A Proposal of Recommendation Agents for Surprising Recipes", Journal of the Institute of Electronics, Information and Communication Engineers D Vol. j98-D No.6 pp.971-981, 2015

[4] Yasuhiro Mori, Eriko Yoshimura, Seiji Tsuchiya, Koichi Watanabe, "Construction of a Cooking Concept Base for Cooking Systems", Information Processing Society of Japan Vol. 2014-ISC-174 No.6, 2014.

[5] Yuka Shidochi, Tomokazu Takahashi, Ichiro Ide, Hiroshi Muras " Discovery of replaceable materials from cooking recipe texts ", The 22nd Annual Conference of the Japanese Society for Artificial Intelligence, 2008

[6] Yusuke Tagawa, Arata Tanaka, Yuya Minami, Daichi Namikawa, Michio Shimomura, Takahira Yamaguchi" Supporting Diet Therapy Based on Japanese Linked Data ", The 27th Annual Conference of the Japanese Society for Artificial Intelligence, 2013

[7] Hidekazu Shiozawa, "A casual recipe search interface considering the priority of ingredients", Transactions of the Institute of Electronics, Information and Communication

Engineers, Vol. J94-A, No. 7, pp. 458-466, © Institute of Electronics, Information and Communication Engineers, 2011

[8] Shihono Karime, Atsushi Fujii, "Recipe Retrieval Considering Nutritional Balance", Proceedings of the 14th Annual Conference of the Association for Natural Language Processing, 2008

[9] Mayumi Ueda, Kazuyuki Ishihara, Yasushi Hirano, Masashi Kajita, Kenji Mase, "A recipe recommendation method that reflects personal tastes based on ingredient usage history", Japan Database Society Letters Vol.6, No.4, 2008

[11] Miki Kimura, “A Case Study for Text Classification by Employing a Weighted Variable TF-IDF”, Meiji University Literature Research Papers No. 48, 2018

[12] Kihara, H., Ueda, M., and Nakajima, S. “A recipe recommendation method that considers the use of surplus ingredients”, DEIM Forum, E3-3, 2011

[13] Mayumi Ueda, Kazuyuki Ishihara, Yasushi Hirano, Masashi Kajita, Kenji Mase, “A recipe recommendation method that reflects personal preferences based on ingredient usage history”, Symposium on Database and Web Information Systems DBWeb2007, 3C-3, 2007

[14] Izumi Toda, Masaru Kakuma, “A computer-based cooking menu system - an expert system for menu creation”, Information Processing Society of Japan SIG Technical Report, No.91-CH-10, pp.1–8, 1991

[15] Hidekazu Shiozawa, “Interactive cooking recipe search by keywords and their priorities”, Interactive Cooking Recipe Search by Keywords and Their Priorities∗wiss2007

[16] Motoki Moritaku, Yuanyuan Wang, “A proposal for a luxurious recipe recommendation method based on metadata analysis of cooking recipes”, DEIM2022 C24-4

[17] Yamamoto Keita, Hinata Nasu, Kazuhiro Kazama "Method of adding metadata to recipes focusing on food arrangements" DEIM2020 P1-5 (day1 p70)